



## King's Research Portal

DOI:

[10.1515/cog-2016-0151](https://doi.org/10.1515/cog-2016-0151)

*Document Version*

Version created as part of publication process; publisher's layout; not normally made publicly available

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Ji, Y., & Hohenstein, J. (2018). English and Chinese children's motion event similarity judgments. *COGNITIVE LINGUISTICS*, 29(1), 45-76. <https://doi.org/10.1515/cog-2016-0151>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

Yinglin Ji\* and Jill Hohenstein

# English and Chinese children's motion event similarity judgments

<https://doi.org/10.1515/cog-2016-0151>

Received December 21, 2016; revised June 27, 2017; accepted August 20, 2017

**Abstract:** This study explores the relationship between language and thought in similarity judgments by testing how monolingual children who speak languages with partial typological differences in motion description (English and Chinese) respond to visual motion event stimuli. Participants were either Chinese- or English-speaking, 3-year-olds, 8-year-olds and adults (32 in each group) who judged the similarity between caused motion scenes in a match-to-sample task. The results suggest, first of all, that the two younger groups of 3-year-olds are predominantly path-oriented, irrespective of language, as evidenced by their significantly longer fixation on path-match videos rather than manner-match videos in a preferential looking scheme. Using categorical measurement of overt choices, older children and adults also showed a shared tendency of being more path-oriented. However, the analysis using continuous measurement of reaction time revealed significant variations in spatial cognition that can be related to linguistic differences: English speakers tended to be more manner-oriented while Chinese speakers were equally manner- and path-oriented. On the whole, our findings indicate a likelihood that children's non-linguistic thought is similar prior to internalising the lexicalisation pattern of motion events in their native languages, but shows divergences after such habitual use, thus suggesting that the pattern of non-linguistic thought may be linked, among other things, to linguistic structure.

**Keywords:** language and space, motion event typology, linguistic relativity, similarity judgment, preferential looking

## 1 Introduction

This study aims to examine linguistic relativity in the domain of space, particularly with regard to motion events. As Papafragou et al. (2002) noted, the spatial

---

**\*Corresponding author: Yinglin Ji**, Research Centre for Language and Cognition, Shenzhen University, Shenzhen City, PR China, E-mail: [yinglin.ji@szu.edu.cn](mailto:yinglin.ji@szu.edu.cn)

**Jill Hohenstein**, School of Education, Communication & Society, King's College London, London, UK, E-mail: [jill.hohenstein@kcl.ac.uk](mailto:jill.hohenstein@kcl.ac.uk)

domain provides an ideal testing ground for the hypothesis of linguistic relativity. First of all, location and motion are fundamental to human life and culture; they involve sensory representations that are biologically determined and shared between humans and other animals. Furthermore, world languages show striking systematic variations in the linguistic encoding of the semantic dimensions of location and motion (i.e., motion event typology proposed by Talmy 1985; Talmy 2000). Equally importantly, language-independent, objective means exist to measure discrimination for spatial perception and conceptualization. Examples include gestures, non-verbal categorisation, recognition memory, similarity judgment and online eye tracking. Such particular characteristics of the spatial domain make the relationship between language and thought a topic that has been increasingly explored by researchers from multiple disciplines. Most studies on space, language and cognition tend to focus on motion events (as compared to situations of static location), in particular, self-instigated voluntary or spontaneous motion events (see, for instance, Allen et al. 2007; Berman and Slobin 1994; Bowerman and Choi 2001; Hickmann 2006; Hohenstein 2005; Slobin 2004). There is a growing body of literature focusing on a more complex type of events involving causality, viz., caused motion events, but most of these studies do not test the linguistic relativity hypothesis directly (see, for instance, Berthele 2015; Cadierno et al. 2016; Filipović 2013; Gullberg 2011; Gullberg and Narasimhan 2010; Ibarretxe-Antuñano 2012; Kopecka and Narasimhan 2012; Wolff 2003). In this context, the present study explores how English and Chinese<sup>1</sup> speakers judge the similarity between caused motion scenes, which encompass a set of particularly rich components such as Path of motion, causing action and different types of Manner information. The complex nature of caused motion events, as illustrated in this study, will help bring out how speakers of different languages weigh the salience of varied types of motion elements in their judgments and to what extent their behavior can be related to language differences.

Languages differ a great deal in how they typically describe motion events. As proposed by Talmy (1985; 2000), some languages (e.g., English and German) tend to lexicalise Manner in the main verb of a motion description, whilst encoding Path in verb-supporting elements such as particles (example 1). In contrast, many other languages (e.g., French and Spanish) tend to incorporate Path into the main verb of a sentence, and express Manner, if necessary, in the periphery of a motion description via gerunds or adverbials (example 2). Due to the placement of the Path information, such differences are summarised by Talmy as satellite-framed (e.g., English) versus verb-framed (e.g., French).

---

<sup>1</sup> The term ‘Chinese’ refers to Mandarin Chinese throughout the paper.

- (1) *He pulled it across the road.*
- (2) *Il traverse la rue en tirant.*  
 he crosses the road while pulling  
 'He is crossing the road while pulling.'

This motion event typology should be viewed as a continuum rather than as an absolute dichotomy (see, for instance, Berthele 2004; Filipović 2007; Goschler and Stefanowitsch 2013; Ibarretxe-Antuñano 2009). There is arguably a third type of language that stands midway along the satellite- versus verb-framed cline, i.e., the equipollently-framed language (e.g., Ameka and Essegbey 2013; Chen 2005; Chen and Guo 2009; Ji et al. 2011c; Slobin 2004; Zlatev and Yangklang 2004). These languages (e.g., Chinese and Thai) are characterised by encoding varied pieces of motion information (such as Manner, Path and Cause) in semantic elements that have equal grammatical status and formal significance: for instance, co-verbs or verb compounds. To illustrate, in Chinese, motion events are typically expressed in a Resultative Verb Compound (RVC), which normally consists of two Constituents [C]: C1 for encoding Manner of motion (e.g., 'walk' in example 3) and C2 for Path of motion (e.g., 'exit'/'out of' in example 3), as illustrated below.

- (3) [C1]–[C2]  
*Ta1 la1 zhe ta1 zou3-guo4 ma3lu4*  
 he pull durative aspect it walk-cross street  
 'He, pulling it, walked across the street.'

As can be seen from the contrast between example 1 and 3, the patterns of motion event description in Chinese and in English are at least partially different. Caused motion is expressed in English in a Manner-and-Cause verb + Path particle combination, while in Chinese, the same event is characteristically encoded in an RVC. Further, caused motion is expressed via a syntactically simplex structure in English whereas in Chinese, it is realised through a complex ZHE construction that is similar to the English *V-ing* form.

These observations are consistent with the status of Chinese in motion event typology, which has been controversial for some time. The initial disagreement centred around the issue of the grammatical category of Path constituent in an RVC (e.g., *guo4* 'cross/across' in *zou3-guo4* 'walk-cross/across'). Talmy (2000) argued that this is a closed-class<sup>2</sup> set of grammatical elements akin to English

<sup>2</sup> As one of our anonymous reviewers pointed out, one could argue that verb-framed languages such as French or Spanish also have a 'closed class' set of Path verbs. Seen this way, the

particles and thus Chinese should be classified as satellite-framed, like English. In contrast, some linguists have pointed out that the syntactic function of C2 in an RVC in Chinese differs considerably from an English particle in that it can stand as an independent predicate (e.g., *guo4* ‘cross’ in *guo4 ma3lu4* ‘cross the street’). Therefore, Chinese and other serial verb languages should be considered in their own right and categorised as ‘equipollently-framed’ (Chen 2005; Chen and Guo 2009; Ji et al. 2011a; Slobin 2004).

A majority of sinologists treat Path constituents in an RVC as verbs rather than particles (see, for instance, Chen 2008; Kang 1999; Li 1990; Li and Thompson 1981; Zou 1994), or at least as a distinctive grammatical category of ‘post-verb’ intermediate between verb and particle (McDonald 1995). Some other linguists (e.g., Tai 2003) have even suggested that Chinese should be taken as verb-framed, based on semantic criteria. According to Tai (2003), for example, the semantic schema for a Chinese RVC is ‘C1 action – C2 result’, where the ‘result’ denotes the attainment or non-fulfilment of the action concerned. Thus, this element is more semantically salient and represents the conflation pattern of Motion + Path, on the basis of which Chinese should be primarily a verb-framed language. Furthermore, an increasing number of recent studies, conducted with different types of motion events (voluntary versus caused), using naturalistic or elicited data, and at the level of semantics or semantic-syntax interface, all suggest that Chinese possesses both satellite- and verb-framing properties (Chen 2005; Chen and Guo 2009; Chu 2009; Gao 2001; Ji et al. 2011a; Ji et al. 2011b; Ji et al. 2011c; Slobin 2004). After examining a complete list of Chinese spatial elements by using an expanded set of main verb criteria, Talmy more recently concluded that Chinese is the only language where the case of equipollence can be found (2009). Given the above views on the typological status of Chinese, it is held in this paper that Chinese is essentially equipollently-framed; and with respect to caused motion expressions in particular, it differs widely from English lexically as well as syntactically.

Some recent studies reveal that language-specific categories can be related to adults’ spatial reasoning at a deeper cognitive level, particularly in voluntary motion events (e.g., Hohenstein 2005; Soroli and Hickmann 2010). In this context, the present study investigates, using a non-verbal similarity judgment task, language-specific influences on mental representation of caused motion events in native Chinese speakers and native English speakers of various ages.

---

classification of Chinese as satellite-framed based on the concept of a ‘closed class’ set of grammatical units seems problematic.

## 2 Motion representation in language and cognition

The effect of typology on language use has been found, in children and adults, in naturalistic and elicited contexts, at multiple linguistic levels, for instance: lexicalisation pattern, syntax, semantics-syntax interface and discourse (Allen et al. 2007; Athanasopoulos and Bylund 2013; Bowerman 1996; Bowerman 1999; Bowerman and Choi 2001; Bylund and Athanasopoulos 2015; Filipović and Jaszczolt 2012; Flecken et al. 2014; Hickmann 2003; Hickmann 2006; Hohenstein et al. 2004; Ji et al. 2011a; 2011b; Pavlenko and Volynsky 2015).

As far as caused motion events are concerned, previous studies explore varied issues such as how the concept of causation is related to the linguistic structure (e.g., direct causation event is more frequently encoded in simplex sentences rather than in periphrastic structures; see Wolff 2003 for details), how the semantics of placement verbs in typologically different languages affects speakers' linguistic construal of placement events (e.g., Cadierno et al. 2016; Gullberg 2011; Gullberg and Narasimham 2010; Kopecka and Narasimhan 2012; to name a few), and how language-specific properties influence the semantic density of caused motion descriptions of children from different languages (e.g., Ji et al. 2011c). Most of these studies suggest that language-specific factors have a larger role to play in event conceptualization than previously assumed. A case in point is Gullberg's (2011) study in which the author investigates whether there are language-specific modes of representation in a presumably cognitively basic or universal domain of placement events (e.g., events of *putting*, *hanging* or *laying*). Her gesture analyses reveal language-specific gesture patterns across the entire placement domain: Dutch (satellite-framed) speakers tend to focus on Figure objects in conjunction with the Path of movement whereas their French (verb-framed) counterparts focus on simple Paths towards Ground (Gullberg 2011: 19–20), a pattern that can be mapped onto the typological difference between the two languages, in particular, the verb semantics (e.g., posture verbs in Dutch vs. a general placement verb 'put' in French).

Similarly, in an investigation using caused motion stimuli consistent with those employed in the present study, Ji et al. (2011c) reveal that speakers express denser semantic information in Chinese than in English because of the availability in Chinese of an easily accessible resultative verb compound which facilitates the simultaneous encoding of varied semantic components for motion. As a result, a striking developmental tendency occurs in English between the ages of 3 to 6 and adulthood, whereas in Chinese such a developmental tendency is significantly less pronounced.

As regards the question of whether language-specific patterns can go beyond language use and affect spatial conceptualization at a deeper level of cognition, some studies have produced evidence in support of linguistic relativity (Berthele 2013; Boroditsky 2000; Boroditsky 2001; Boroditsky and Gaby 2010; Blomberg and Zlatev 2009; Bohnermeyer et al. 2004; Finkbeiner et al. 2002; Hohenstein 2005; Levinson 2003; Lupyan 2012; Pederson et al. 1998; Wolff and Holmes 2011; Zlatev 2011; Zlatev and Blomberg 2015). For instance, Levinson (2003) noted that spatial thinking can co-vary with languages: there were robust correlations between frames of reference used in language (e.g., intrinsic, relative or absolute) and frames of reference used in non-linguistic memory and reasoning, suggesting a Whorfian effect of languages on cognition. With respect to motion events, a series of studies by Bowerman and Choi (Bowerman 1996; Bowerman 1999; Bowerman and Choi 2001; Choi and Bowerman 1991) revealed that there was a pervasive interaction between non-linguistic conceptual development and the relevant semantic category of an input language. English Path markers, typically prepositions, do not reflect whether a motion is presented as spontaneous or caused, while Korean marks most Path meanings with distinct verb roots in transitive and intransitive clauses. Accordingly, English-speaking children used the prepositions *up* and *down* for a great range of spontaneous and caused motions along the vertical axis, including climbing on and off furniture and raising and lowering things, which suggested that they possessed a core spatial concept of vertical directionality. By contrast, Korean children consistently distinguished different kinds of vertical motion and used different verbs (transitive versus intransitive) for spontaneous and caused motion along a Path.

Previous studies exploring the subtle links that languages may have with cognition have used various non-linguistic tasks. For example, Hohenstein (2005) incorporated a *preferential looking* paradigm with a temporal link between spatial language acquisition and cognitive development. She examined monolingual Spanish- and English-speaking children's responses to visual voluntary motion stimuli and found that in a match-to-sample task, participants behaved differently towards video stimuli in ways that could be predicted by their respective languages: 7-year-old English-speaking children fixated on videos matching the Manner (rather than Path) of a target video for longer than did Spanish-speaking 7-year-olds or either group of 3.5-year-olds.

Other studies examining non-verbal performances have reported a less direct effect of language on cognition. For instance, Gennari et al. (2002) analysed participants' performances in non-linguistic tasks in order to explore whether different lexicalisation patterns in English and Spanish have a cognitive significance for the speakers of these languages. A linguistic effect was reported in the similarity judgment task, where participants were asked to make their

choices after they had been encouraged to verbally describe relevant voluntary and caused motion events. This suggested that the language-specific regularities that are available in the experimental context can help mediate participants' performance in non-linguistic tasks (Gennari et al. 2002). In a similar fashion, Soroli and Hickmann (2010) examined whether the differences in lexicalisation patterns of voluntary motion events in English and French affected speakers' non-verbal behaviour, and if so, how. They reported that French speakers preferred Path to Manner as their categorisation criterion in the verbal as well as the non-verbal categorisation tasks. English participants showed no preferences for Path or Manner in the non-verbal categorisation task, but showed a significant preference for Manner in the verbal categorisation task, where the linguistic encoding of a given motion event was provided before participants were asked to categorise it. These findings provide evidence in support of a weak version of the relativity hypothesis: that is, language effects on non-verbal behaviour exist solely when relevant linguistic information is simultaneously provided. Such findings are consistent with Slobin's 'thinking for speaking' hypothesis, suggesting that "the activity of thinking takes on a particular quality when it is employed in the activity of speaking" (1996: 75–76).

There are also studies that reported that language had no (or very unclear) effects on spatial cognition (Durst-Andersen 2011; Jackendoff 1996; Landau and Lakusta 2006; Lucy 1992; Mandler 1992; McWhorter 2014; Papafragou et al. 2002; Spelke 2003). Papafragou et al. (2002) compared the performance of English and Greek children and adults in memory and categorisation of motion events. Their findings indicated that there was no language effect, despite the fact that speakers of the two languages differed greatly in terms of their verbal descriptions of the motion events.

Different results obtained from various studies may stem from divergences in methods such as the nature of the stimuli used (static pictures/photographs versus dynamic video clips); the presence or absence of 'shadowing' conditions; the existence or absence of a time limit in viewing stimuli and making choices; as well as different means of measuring cognition (categorical preferences versus visual fixation). To illustrate, although Papafragou et al. (2002) and Hohenstein (2005) experimented with speakers of satellite- (English) versus verb-framed (Greek and Spanish, respectively) languages,<sup>3</sup> they arrived at different conclusions about language effects on spatial cognition. This could result from, among other things, differences in methods: in the former study,

---

<sup>3</sup> Note that the exact status of Greek in motion event typology is a debated topic. Some scholars argue that it is not typically verb-framed, but shows satellite-framed properties or has a 'parallel' or 'mixed' system in encoding motion events (see, for instance, Skopeteas 2008; Talmy 2009).



participants were given unlimited time to view the static stimuli and make their choices, whereas the latter investigation involved strict time constraints and dynamic stimuli. Strictly speaking, Papafragou et al.'s (2002) study is not a truly 'nonlinguistic' one in the sense that the lack of time constraints may have allowed the speakers to use language as a strategic tool in solving the problem, therefore, a subtle influence of language on conceptualization can be expected. However, no such language effect was attested. As Hohenstein (2005) suggests, differences in non-linguistic thought resulting from linguistic variations are transient in nature and may not last. Thus, given an unlimited amount of time to study the stimuli, speakers of different languages would have made the same choice. This may partly explain why speakers of English and Greek (satellite-framed vs. verb-framed) behave similarly in judging the similarity of static motion scenes in Papafragou et al.'s (2002) study.

It is also worth noting that previous investigations of spatial conceptualization have examined groups of speakers of languages with opposing typological properties (i.e., satellite- versus verb-framed). Only a very small number of investigations have systematically examined spatial cognition with respect to languages that show less dramatic differences (e.g., equipollently-framed Chinese versus satellite-framed English). In fact, there is an increasing number of studies suggesting that the 'intra-category' variation in terms of framing properties of a given language is much more common than is previously assumed. For instance, some traditional verb-framed languages such as French or Italian are found to have clear satellite-framed properties. Kopecka (2006) points out that the Path component is not only expressed in the verb in French, but also in a prefix showing a satellite-framed pattern with high degree of productivity (2006: 85–91). In a similar fashion, apart from conflating Path and Motion in the verb root, Italian has a group of post-verbal particles that add directional values to the verb root and function as true satellites (Iacobini and Masini 2006: 161). There are even frequently used colloquial Path verbs such as *rise*, *fall* and *sink* (apart from those Path verbs of Latinate origin such as *ascend*, *descend*, *enter*, *exit*) in English, a language traditionally and standardly considered to be satellite-framed (Beavers et al. 2010). The widely attested variations in terms of framing properties within a language of a given typological group suggest that the Talmyan bipartite typology is better seen as representing tendencies rather than absolute differences between languages. Given this context, the present study aims to investigate the potential influence of typological tendencies of languages on the thought pattern of speakers. Specifically, if mental representations of motion events differ between Chinese and English speakers, such findings would constitute evidence in support of the relativity hypothesis: subtle differences between languages, rather than categorical variations, would be shown to influence non-linguistic cognition.

Furthermore, many of the studies examining the relationship between language and thought have been conducted with adults only (e.g., Gennari et al. 2002; Levinson 2003; Soroli and Hickmann 2010). One criticism of such studies is that differences between the groups studied could be due to culture rather than language. A better test of linguistic influence on cognition would involve children because this would more closely tie language acquisition to cognitive development (e.g., Hohenstein 2005). That is, if children's cognition is similar prior to becoming accustomed to using the spatial language in ways typical to their native language, but then shows differences after such habitual use, the argument that this difference occurs due to 'general culture' becomes weaker.

In this context, the current study conducted a similarity judgment experiment with the aim of investigating the relationship between language and spatial conceptualization with particular reference to languages beyond the satellite- versus verb-framed dichotomy (i.e., Chinese and English). Although previous research has shown that there is some link between children's cognition and language acquisition in verb- and satellite-framed languages (Hohenstein 2005), it remains to be seen whether similar links would be evident in young learners of an equipollent language like Chinese.

### 3 Motion description from a developmental perspective

Some studies involving children use continuous measures of cognition for motion events, which typically include eye tracking and preferential looking (e.g., Dittmar et al. 2008; Flecken 2011; Papafragou et al. 2008; Pulverman et al. 2013; Von Stutterheim et al. 2012). Most of them showed a clear influence of language-specificity during the conceptualization phase in language production and event perception. As a case in point, Von Stutterheim et al. (2012) explore how grammaticized concepts shape motion event conceptualization using eye tracking data collected from children of languages as various as Arabic, Czech, Dutch, English, German, Russian and Spanish. Their findings are that the direction of children's attention to particular parts of motion events varies systematically with the existence of grammaticized means to express imperfective/progressive aspect in one's native language. In a few studies where a direct language-specific influence on cognition is not reported, the results still suggest that as long as some kind of relevant linguistic experience is recruited immediately prior to testing, children's behaviour shows cross-language differences in

motion event perception and early understanding of syntactic marking (e.g., Dittmar et al. 2008; Papafragou et al. 2008).

Children in this study were of two different ages (3 years and 8 years), based on the expectation that patterns of language acquisition could be linked with regularities in non-linguistic spatial cognition. Specifically, these two age groups represent points in development of Manner-Path use in children's language, particularly the generalisation of motion verbs, as revealed in previous studies (e.g., Choi and Bowerman 1991; Hohenstein 2005; Ji et al. 2011a; 2011b).

Although children as young as 3 years of age are reported to be generally sensitive to the method of expressing motion that is characteristic of their native language, their mastery of motion-related language is far from stable. Specifically, although English-speaking children start combining Path particles with a variety of verbs specifying Manner and Cause from about age 2 (Berman and Slobin 1994; Choi and Bowerman 1991), their productivity at this stage is still limited. Bowerman (1982; 1988), for example, reported that it is only about a year and a half later (i.e., around the age of 3;6) that English-speaking children fully acquire the characteristic way of encoding motion events, as evidenced by their production of most conventional combinations of Manner verbs and Path particles in a systematic way, and also by their ability to create combinations of verbs and Path particles with novel meanings. Similarly, although Chinese children are reported to be able to use RVCs frequently and productively for their motion expression from an early age of 2;6, it is not until the age of 3;6 that the children learn to assign different argument structures to different types of RVCs. Their knowledge of the transitivity of constituent verbs and the semantics of the Path constituent still requires further development (Chen 2008).

By the age of 8 years, children are found to have fully acquired the Manner-Path use that is characteristic of their native language. For example, Ji et al. (2011a; 2011b) investigated how English and Chinese children of different ages verbally describe voluntary and caused motion events. Their findings indicated that the 8-year-olds' verbal behaviour mirrors that of adults, both quantitatively (e.g., in terms of the number of motion elements encoded in an utterance) and qualitatively (e.g., in terms of making fine semantic distinctions in the expression of Manner information and in the accumulation of Path elements). In the use of motion verbs, the 8-year-olds acted like adults, using Manner verbs in English and verb compounds in Chinese, respectively, in a stable and a typical way. As such, this study aims to test whether the pattern of motion language acquisition, as demonstrated above, can manifest itself in the pattern of non-linguistic thought in children.

## 4 Predictions

As illustrated earlier, one important difference between English and Chinese in motion description is that the grammatically marked category of verb tends to encode only Manner in English, but Manner and Path simultaneously in Chinese. Because the verb is arguably more central to an utterance, the English speakers' attention should be drawn more to Manner, whereas Chinese speakers are likely to pay an equal amount of attention to Manner and Path. If such kinds of language differences can foster different patterns in spatial thinking in non-linguistic tasks, we expect English speakers to be more oriented towards the Manner of motion in the similarity judgment than Chinese speakers. Meanwhile, it is argued in some previous studies that if an aspect of language is purported to influence how people perceive events, children's non-linguistic performance should look similar across languages before they acquire the linguistic differences in question and they should develop related cognitive differences only afterwards (see Hohenstein 2005).

As such, it was predicted in the current study that the 8-year-old English children, like English adults, would tend to be more Manner-oriented and choose Manner-matched motion scenes in the similarity judgment more frequently than would the Chinese older children and adults, as well as the 3-year-olds in both language groups. In contrast, 8-year-old Chinese children as well as Chinese adults are expected to prefer neither Manner-match nor Path-match scenes. Furthermore, because the younger children of 3 years do not yet have a firm grasp of the lexicalisation pattern of their language, and also because there is assumed to be a universal cognitive salience of Path in a motion event (Talmy 2000), the 3-year-olds are predicted to be more Path-oriented, irrespective of their language; and their frequency of preferring Path-matched motion scenes is expected not to differ significantly between language groups. A number-shadowing (repeat aloud) condition was used in the match-to-sample-task to prevent 8-year-olds and adults from verbalising motion events subconsciously, thus increasing the likelihood that the present study was a truly 'non-linguistic' one.

## 5 Method

### 5.1 Participants

Participants were monolingual native speakers of English and Chinese who fell into six groups. Two adult groups comprised university students from London

and Beijing, respectively. Four groups of children, aged 3 years and 8 years, were recruited through primary schools and day nurseries in London and Beijing. A total of 192 people participated in the study, with 32 participants per group (16 males and 16 females). Detailed information about the participants can be found in Table 1.

**Table 1:** Groups of participants in the study.

Group ID	Native language	Age level	Number of participants	Mean age
C03	Chinese	3 years	32	3;2 ( $SD = 0.31$ )
C08	Chinese	8 years	32	8;2 ( $SD = 0.02$ )
CAD	Chinese	Adults	32	19;3 ( $SD = 0.97$ )
E03	English	3 years	32	3;2 ( $SD = 0.32$ )
E08	English	8 years	32	8;3 ( $SD = 0.03$ )
EAD	English	Adults	32	26;0 ( $SD = 5.17$ )

Note: The difference in the mean age between Chinese and English adults should not matter; both groups of participants were generally considered to have reached the ‘end-state’ of their spatial language use and development of spatial cognition.

## 5.2 Materials

The stimuli consisted of short video clips, comprising a set of 48 motion events (Ji and Hohenstein 2017c). These video clips were organised as a set of 16 triads: 16 targets and 32 alternates (two for each of the target events). Target and alternate videos appeared for 5 sec each and were then followed by 1 sec of black screen. The task lasted approximately 6 min total.

Within a triad, the target video clip depicted a boy performing a specific action which caused the movement of an object; at the same time, the boy accompanied this object during the course of its movement (e.g., the boy dragged a barrel around the table) while the two alternate video clips showed the same boy involved in similar actions with variations in either the Manner or the Path dimension. In the Path-match alternate, the Path of motion was preserved while the Manner of motion was changed (e.g., the boy *rolled* a barrel around the table), while in the manner-match alternate, the Path was altered while Manner was held constant (e.g., the boy dragged a barrel *away from* the table). To maximise similarity, all stimuli involved the same protagonist with the same clothing. Within each triad, the object in motion, apart from the boy, remained the same and the surroundings in the target and alternate scenes were kept uniform. A complete list of stimuli, along with an example illustration, is given in Appendix A and B.

Our stimuli demonstrated motion events of a particular type, that is, caused motion involving accompanied movement on the part of the protagonist. This design followed the model of caused motion developed by Hickmann et al. (2009) with the following properties: a) each stimulus involved a boy and an object, both shifting locations from one point to another in motion; b) the object's motion was provoked by an external force exerted on it by the boy; and c) the boy's continued action was necessary for the object's motion and the boy and the object followed the same trajectory in the course of their movement.

All stimuli were arranged into two randomised orders, the second being the reverse of the first. These orders were counterbalanced across participants within each language and age group. The presentation position of Manner-match versus Path-match video clips (left or right of the screen) was counterbalanced across stimuli in a given order.

## 5.3 Procedures

### 5.3.1 Pretest

For 8-year-olds and adults, a practice item preceded the task (a triad of video clips showing a male character engaging in specific causing actions with respect to certain Paths). This was intended to familiarise participants with the test procedure and to direct their attention to the fact that the matching judgment was expected to be made based on the similarity of actions. For 3-year-olds, there was a pretest for familiarisation and perseveration. The 3-year-olds were shown three triads of static pictures showing pigeons, flowers and bananas, respectively. Within each triad, the target object was placed in the centre of a piece of paper while two alternates were shown side-by-side on a separate page, each differing from the target by either size only or colour only. The children were then asked to identify which of the two alternates was most like the target object. This mirrored the task they were expected to perform in the testing session. If children had chosen alternates from only one side of the page for all three selections (none did), they would have been considered perseverative and eliminated from the study.

### 5.3.2 Testing session

Participants were tested individually by a single female experimenter. Groups of adults and older children (8 years) were invited to watch video clips displayed on a MacBook Pro and asked to signal their judgments of similarity between motion scenes by pressing one of two keys on the keyboard, these being letters

‘A’ and ‘L’ respectively, which were kept wide apart from each other on the keyboard and covered with white stickers (i.e., no linguistic labelling). 16 triads of video clips were played to these participants on the laptop screen through the stimulus presentation software ‘SuperLab 4.5’, which generated a file with recorded information about the participant’s choices (Manner-match or Path-match) and his or her reaction time (RT). Demographic information was collected from participants prior to their testing session. Written instructions and consent procedures were presented to participants in their native language.

The video set-up was the same in all test sites. The stimuli were played in a synchronised series with the target videos playing first in the centre of a screen with a black background, followed by two simultaneous alternate videos placed side-by-side on the same screen. A black screen was shown for 0.5 sec between a target and two alternates within each triad, and for 1 sec between the triads. The participant was instructed to watch each video clip and to listen to the directing audio that first labelled each target with a number and then directed the participant to judge, in his or her opinion, which of the two alternates looked most like the target (e.g., “This is 1. Which one is most like 1?”). Participants were reminded to make their choices as quickly as they could. Adults were asked to count backwards from 100 to 1 and repeat this process several times as appropriate whilst they watched and made their judgments on the stimuli (the children of 8 years were asked to count forwards from 1 to 100). The purpose of this shadowing task was to prevent participants from subconsciously verbalising motion scenes during the testing session.

For the two groups of younger children of 3 years of age, the preferential looking technique was adopted. The stimuli played via SuperLab on the laptop screen were projected onto a 16: 9 format projection screen (120 cm in length and 68 cm in width). A video camcorder was placed just above the screen in a central position to record the younger children’s eye movements whilst they watched the video clips. Each 3-year-old sat on a low stool approximately 1 m away from the projection screen with the teacher of his or her class sitting behind and to the right to provide reassurance to the child. In particular, the teachers were reminded that they must not watch the videos with the children because this may influence the children’s responses in a subtle way that themselves may not be aware of. As soon as video clips were started, the female experimenter removed herself from the child’s view by retreating to a corner of the testing room.

## 5.4 Coding

Two types of variables were used to measure the data: a. the categorical variable, these being overt choices (Manner-match or Path-match) made by older

children and adults; b. the continuous variable, which was further categorised into RT (in milliseconds) of older children and adults and fixation time (in sec) of the 3-year-olds.

In the match-to-sample task, judgments of adults and older children were classified as preferring either the Manner-match or the Path-match according to the button that was selected. The RT for a given stimulus was calculated from the time between the onset of presentation of alternate videos in a triad until their completion, with a theoretically longest RT of 6,000 milliseconds (ms). The RT data was first cleaned by excluding physically impossibly short values (button pressed within 200 ms of stimulus onset). As for extremely long values, screening for outliers was conducted by removing all observations that were more than two standard deviations from the mean.

In the preferential looking experiment, the children's eye movements were recorded for later coding of fixation to Manner-match and Path-match screens. If children fixate more on the Manner-match video than the Path-match video, then it follows that they are more Manner-oriented in similarity judgments of motion events. The continuous measure of visual fixation focused on differences in attention to motion rather than overt choice differences between language groups. Providing a means of determining to what extent children of different language groups view Manner and Path as salient may better reveal subtle differences in behaviour resulting from relatively small language differences. Specifically, video recordings of younger children's eye movements were coded in a frame-by-frame mode using QuickTime Player 7.6. The measure used in the analysis was the number of frames the child fixated on the Manner-match minus that they fixated on the Path-match. Twenty percent of the video data (12 out of 64 files) were recoded by a second coder. *Kappa* indicated that the interrater reliability was ideal ( $\kappa=1.0$ ).

## 6 Results<sup>4</sup>

### 6.1 Preferences for the Manner-match (or the Path-match) across 6 participant groups

Our first set of analyses tested the prediction that the English 8-year-olds and adults would tend to be more Manner-oriented, choosing the Manner-match

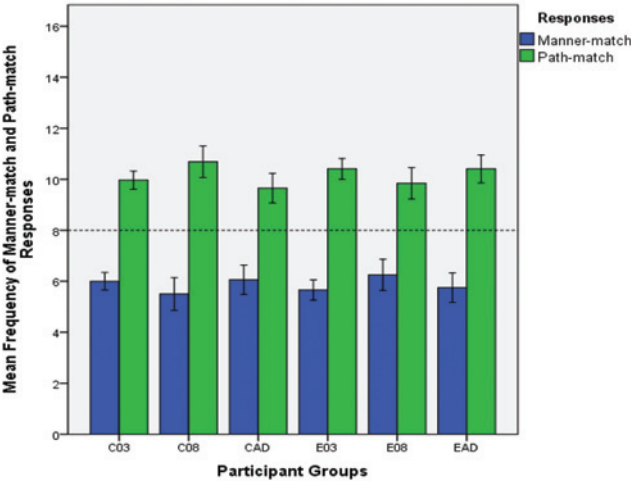
---

<sup>4</sup> Replication data for all the results reported in this section are available via doi:10.18710/AAZVJH (Ji and Hohenstein 2017a).



videos more frequently than age-matched Chinese participants, and more than the 3-year-olds in both language groups. The data of categorical preferences were represented as falling into one of the two major strategies in judging the similarity between screens: the Manner-match and the Path-match. In older children and adults, preferences were decided according to which button they pushed while viewing. As for the younger children, their preferences were determined by difference scores between the number of frames they fixated to the Manner-match minus that to the Path-match for each item. Thus, positive numbers represented a preference for the Manner-match and negative numbers indicated a preference for the Path-match. The mean was then calculated by recording the number of matches out of 16 across individuals in a group.

Figure 1, below, presents the mean number of both Manner-match and Path-match preferences across 6 participant groups (C03, C08, CAD, E03, E08 and EAD). There appeared to be a shared pattern of preferring the Path-match over the Manner-match (see also Table 2).



**Figure 1:** Mean number of Manner-match and Path-match preferences across participant groups.

Given that a visual inspection of Figure 1 suggested a shard preference for the Path-match across participant groups, a set of chance analyses was performed, in the first instance, to determine whether the Path-match was the predominat-ing decision strategy in each individual group. The results corroborated our observation: the number of Path-match preferences fell above chance levels (test value = 8) across all participant groups (more Path-matches): Chinese

**Table 2:** Means and standard deviations (SDs) for the number of Manner- and Path-match across participant groups.

Language	Age	Means (SDs) for Manner-match	Means (SDs) for Path-match
Chinese	3 years	6.00 (1.951)	10.00 (1.951)
	8 years	5.50 (3.646)	10.50 (3.646)
	Adults	6.06 (3.252)	9.94 (3.252)
	Total	5.85 (3.016)	10.15 (3.016)
English	3 years	5.66 (2.252)	10.34 (2.252)
	8 years	6.25 (3.455)	9.75 (3.455)
	Adults	5.75 (3.292)	10.25 (3.292)
	Total	5.89 (3.026)	10.11 (3.026)

3-year-olds:  $t(31) = 5.799, p < 0.001$ ; Chinese 8-year-olds:  $t(31) = 3.879, p < 0.01$ ; Chinese adults:  $t(31) = 3.370, p < 0.01$ ; English 3-year-olds:  $t(31) = 5.887, p < 0.001$ ; English 8-year-olds:  $t(31) = 2.865, p < 0.01$  and English adults:  $t(31) = 3.866, p < 0.01$ , respectively.

Further, in order to test whether the number of the Path-match preference will vary with language and/or age, logistic mixed effects analyses were carried out using R (version 3.4.0) with glmer function in the lme4 package. The Path-match choice was used as the nominal dependent variable. Language, age group and their two-way interaction were entered into the model as fixed effects. Language was coded as a 2-level factor (Chinese, English) and age group was coded as a 3-level factor (3 years, 8 years and adults). Random effects included random intercepts for participants and items, and by-item random slopes for language and age group. The analyses revealed no main effect of either language or age, nor any significant interaction effect between the two variables. Regression coefficients for the model were presented in Table 3. These results did not provide support to our prediction that the English-speaking older children and adults would be more Manner-oriented. As indicated in Figure 1, the frequency of Path-match choices was comparable not only between language groups, but also across age levels.<sup>5</sup>

<sup>5</sup> An additional analysis was conducted to see whether the number of the Path-match remained systematic across individual items. A repeated measure analysis of variance (ANOVA) with language (Chinese, English) and age levels (3 years, 8 years and adults) as between subjects factors and item (16) as a within-subjects factor showed a significant difference across items,  $F(15, 2790) = 14.202, p < 0.001$ , suggesting a likelihood that a particular item was judged as more salient in Manner or in Path. For example, the frequency of the Path-match with item 12 (target: pulling tyre towards rails; alternates: pulling tyre *along* rails and *throwing* tyre towards rails) fell below chance levels (i.e., more Manner-matches):  $t(191) = -4.379, p < 0.001 (M = 0.35,$

**Table 3:** Regression coefficients for the logistic regression mixed model for Path-matches.

	Estimated parameter	SE of estimate	z-value	Pr
Fixed effects				
Intercept	0.627807	0.269077	2.333	0.0196*
Language	0.112081	0.284713	0.394	0.6938
Age (8 years)	0.147524	0.516964	0.285	0.7754
Age (adults)	0.027784	0.552647	0.050	0.9599
Language : age(8 years)	−0.428631	0.382961	−1.119	0.2630
Language : age (adults)	−0.009235	0.384268	−0.024	0.9808
Random effects				
	Variance	Standard deviations		
Participant (intercept)	0.8159	0.9033		
Item (intercept)	0.5867	0.7660		
Language	0.1522	0.3901		
Age (8 years)	3.0981	1.7601		
Age (adults)	3.7091	1.9259		

Note: \* $p < 0.05$

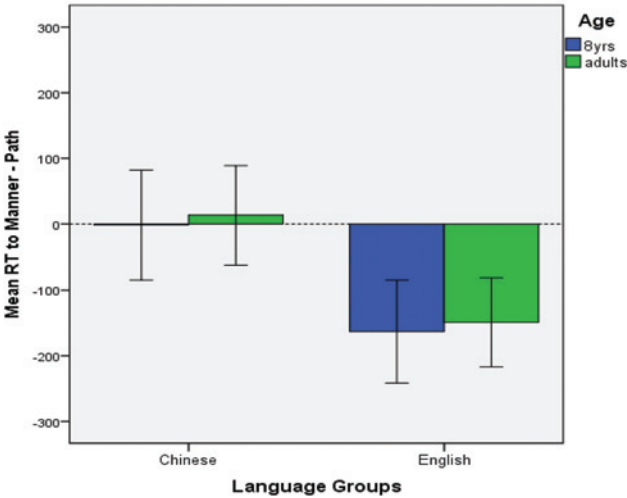
## 6.2 Degree of differences between language and/or age as measured by RT

As previously mentioned, apart from overt choices, the degree of differences between language groups and/or across age levels was also explored by measuring RT (in ms) of the 8-year-olds and adults. Following typological differences between the two languages involved (i.e., satellite-framed English vs. equipollently-framed Chinese) and the presupposed psychological implications (more habitual attention to Manner in English vs. equal amount of habitual attention to Manner and Path in Chinese), English speakers can be arguably more ‘familiar’ with the Manner dimension (vs. Path dimension) in their conceptualization of motion events. Some previous studies exploring children’s language processing using RT measures have revealed that in spoken word recognition experiments, young children responded significantly more quickly in identifying familiar words and matching them with the appropriate referent (see Fernald et al. 2008 for an example). In our paradigm, RT, as response latency, has been used to explore whether and how linguistic factors can

$SD = 0.478$ ), indicating that aspects of the caused motion event as illustrated in item 12 encouraged speakers to make their decisions based on the similarity in the Manner (vs. Path) dimension. A detailed examination of this cross-item difference is beyond the scope of this paper, and the topic will be explored further in our future research.

influence the speed of behavioral decision. If familiarity can indeed enhance speakers’ processing efficiency in decision-making, we then expect that for English speakers, the RT in choosing the Manner-match should be significantly shorter than that in choosing the Path-match, whereas in Chinese, the RT in choosing the Manner-match should be comparable to that in choosing the Path-match.

Thus, our second set of analyses tested whether the difference scores between the RT in Manner-match choices minus that in Path-match choices vary with language and/or age. Positive numbers represented longer RT in making Manner-match responses while negative numbers signified longer RT in making Path-match responses. Prior to conducting analyses, 8 outliers in trials (i.e., those observations that were more than two standard deviations from the mean) were excluded form the dataset. Then, a two-way factorial ANOVA was conducted with language (Chinese, English) and age (8 years, adults) as two between-subjects variables, with RT differences as the dependent variable. It revealed a main effect of language only  $F(1, 116) = 4.525$ ,  $p = 0.036$ ,  $\eta_p^2 = 0.038$ , Figure 2<sup>6</sup>; no significant effects of other types were detected (see also Table 4).



**Figure 2:** Means of RT in Path-match responses subtracted from RT in Manner-match responses by language and age (in number of milliseconds).

6 The effect size based on differences between means (i.e., Cohen’s  $d$ ) increased to 0.40.

These findings indicate that the absolute values of difference scores were significantly smaller in Chinese ( $M = 6.57$ ,  $SD = 429.820$ ) than in English ( $M = -155.90$ ,  $SD = 396.697$  Cohen's  $d = 0.40$ ), thus suggesting a tendency for Chinese participants, irrespective of age group, to use an approximately equal amount of time judging the Manner-similarity and the Path-similarity of screens, whereas their English counterparts reacted significantly quicker in judging Manner-similarity compared with judging Path-similarity.

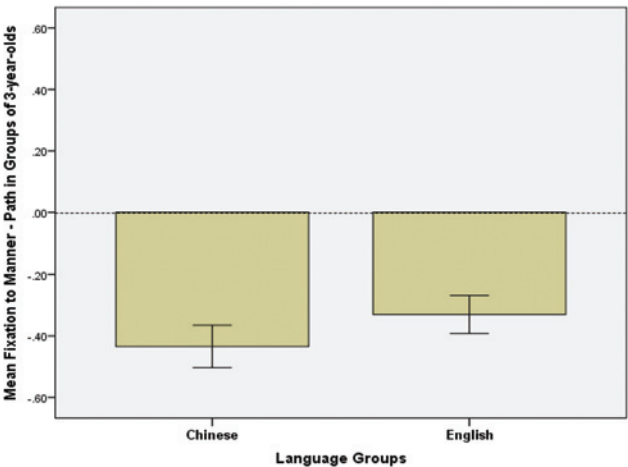
A by-item analysis was further conducted to investigate how systematic the above-reported effect is across items. The measure used was the difference scores between RT of speakers in Manner-matched judgments minus that in Path-matched decisions for each individual item. A repeated measures ANOVA test with language (Chinese, English) and age group (8 years, adults) as within-items factors showed that the effect of language was marginally significant:  $F(1, 15) = 4.314$ ,  $p = 0.055$ ,  $\eta_p^2 = 0.223$ . No simple effect of age or any interaction effect between language and age was detected. Generally speaking, the differences between RT for selecting Manner- and Path-matched scenes were significantly smaller for Chinese participants across age groups ( $M = 73.095$ ,  $SD = 582.95$ ) than for English 8-year-olds and adults ( $M = -250.93$ ,  $SD = 602.382$ ). Thus, our RT analyses by subject and by item, respectively, roughly converged on differences arising between language groups. These results largely map onto the differences in how speakers of Chinese and English linguistically describe and habitually attend to semantic components for motion (Ji et al. 2011c).

### 6.3 3-year-olds' motion event looking time

As revealed earlier in Section 6.1, the younger children of 3 years, irrespective of language, preferred the Path-match over the Manner-match. Such a categorical preference only indicated the category that received more points of fixation; it did not give a degree of difference like the focus on number of frames' difference, which equated to time. Therefore, in this sub-section, we measure the same set of data in terms of fixation time. An independent  $t$ -test was performed to investigate whether the difference scores between the fixation to the Manner-match minus that to the Path-match varied between language groups (cf. Figure 3; see also Table 5). The results suggested that such differences in fixation time were trivial ( $t(62) = -1.127$ ,  $n.s.$ ). Thus, our analyses of younger children's behaviour, using categorical measure (Manner- or Path-match) and continuous measure (difference scores in fixation time), converged on a strong resemblance between Chinese and English 3-year-olds in being predominantly Path-oriented.

**Table 4:** RT (in ms) in Manner-match and Path-match responses across participant groups.

Language	Age	Mean RT in Manner-match	Mean RT in Path-match	No. of people
Chinese	8 years	2819.754 (SD = 631.444)	2760.695 (SD = 627.458)	32
	Adults	2558.160 (SD = 672.789)	2544.734 (SD = 584.407)	32
	Total	2688.957 (SD = 660.534)	2652.714 (SD = 611.250)	64
English	8 years	2885.955 (SD = 732.769)	3003.999 (SD = 666.615)	32
	Adults	1916.054 (SD = 472.128)	2090.603 (SD = 471.137)	32
	Total	2401.005 (SD = 782.820)	2547.301 (SD = 734.689)	64



**Figure 3:** Means of younger children’s fixation to the Path-match subtracted from their fixation to the Manner-match by language group (in number of seconds).

## 7 Discussion

This study tests the theory of linguistic relativity by investigating whether the properties of two languages with relatively small typological differences (English vs. Chinese) influence how speakers perceive and evaluate the similarity between motion scenes on the basis of video clips illustrating a particular type

**Table 5:** Means and standard deviations for fixation time to Manner-match and Path-match screens between Chinese and English 3-year-olds (in number of seconds).

Language	Fixation type	Mean fixation time	Std. Deviation	No. of children
Chinese 3-year-olds	Fixation to Manner-match	33.196	3.502	32
	Fixation to Path-match	40.143	3.649	32
	<i>Total</i>	36.670	3.420	64
English 3-year-olds	Fixation to Manner-match	36.026	2.715	32
	Fixation to Path-match	41.307	3.363	32
	<i>Total</i>	38.667	3.529	64

of motion events: caused motion. As introduced in the opening section, English tends to be satellite-framed: incorporating Manner in the main verb and Path in particles; whereas Chinese is increasingly considered to be equipollently-framed in simultaneously encoding both Manner and Path in a compound verb. Although the co-occurrence of Manner and Path (realised through a verb + particle combination) is also very frequent in English utterances, it is the verb that carries the Manner information, and because the verb is arguably the most central grammatical category, it is likely that the English speakers' attention would be drawn more to the Manner than to the Path. In contrast, the Chinese speakers' attention should be directed equally to the Manner and to the Path. Therefore, we have hypothesized that in a non-linguistic task of similarity judgment, mature English speakers would be more oriented towards Manner of motion while mature Chinese speakers would be equally Manner- and Path-oriented. Younger children at the age of 3, however, may not show any language-biased differences in their similarity judgments given that they have not yet internalised the lexicalisation pattern of motion in their languages.

Speakers' cognitive tendency of being more oriented towards Manner (or Path) is measured in this study by using two variables: the categorical variable of overt choices (i.e., Manner-match or Path-match), and the continuous variable of RT in Manner (or Path)-match choices (for older children of 8 years and adults) and the fixation time in preferential looking (for the younger children of 3 years). Our main findings can be summarised as follows:

- i. As predicted, the performance of the two younger groups of 3-year-olds was remarkably similar in being predominantly Path-oriented, regardless of means of measurement adopted (i.e., categorical preferences, continuous variable of difference scores in fixation time).
- ii. Using the categorical measurement of overt choices, speakers, irrespective of language and age, showed a common tendency of orienting towards

the Path-match. Item analyses revealed that they chose Path above chance levels with most events.

- iii. However, when the continuous measurement was opted for (i.e., difference scores in RT), English participants were significantly quicker in making Manner-matched judgments as compared to Path-matched decisions whereas their Chinese counterparts tended to use an equal amount of time in making Manner- or Path-matched judgments. A further by-item analysis suggested that such a difference in response latency seemed to be largely systematic across items, indicating that RT differences in behavioural decision may not be superficial, but reflecting, at least partially, an effect of relevant language differences (i.e., Manner-salience in English vs. Manner-and-Path salience in Chinese) on cognition for motion.

The results regarding the younger groups of children (i.e., finding i.) can be explained in different ways. Following Talmy's 'Basic Scheme for Motion' (1985), Path information is the most central and indispensable dimension for any motion event. Before children fully acquire the characteristic pattern of motion expressions in their respective languages, the cognitive saliency of Path presents itself in children's non-linguistic behaviour. Further, as regards the particular type of caused motion under discussion, Path information, as compared to Manner, seems to be a more straightforward dimension to follow in viewing. Caused motion stimuli as presented in our experiment involved three specific types of Manner: a) Manner of the protagonist (i.e., *walking*); b) Manner of Cause (i.e., *kicking, throwing, pulling, pushing*); and c) Manner of the object (i.e., *rolling, sliding*). In contrast, although there are two types of Path in each stimulus (i.e., the Path of the protagonist and the Path of the object), these two trajectories are congruent, thus potentially presenting as a more convenient dimension for younger children to stare at.

The divergence between finding ii. and iii. stems, among other things, from different means of measurement (categorical vs. continuous). The categorical preference shows fairly overt choices, which might not be powerful enough to allow potential subtle variations in cognition resulting from linguistic differences to unfold. In contrast, the time of reaction in choosing the Manner (or Path)-match is a continuous variable that might better illuminate the degree of differences in attention to Manner and/or Path between speakers from different languages. This is particularly important for language groups showing only partial differences, such as English and Chinese, rather than fully opposing typological features as in previous studies.

Questions may arise concerning how linguistic differences exert influences on non-linguistic thought patterns. The associationist view holds that there may



be an association between the frequency of verbs and non-linguistic thought. The number of Manner (or Path) verbs used in one's language may command an overarching 'presence' of Manner (or Path) in one's thought pattern (Hohenstein 2005). Slobin (1996) argues that each language provides a specific set of grammatical morphemes for schematising motion experience. "It is through listening that children's attention is first drawn to the fact that certain notions are grammatically marked in the ambient language...and the form in which one receives information from others influences how that information is understood, stored and later accessed"; a native language we learn in childhood "is not a neutral coding system of an objective reality", but instead is a system "that has trained its speakers from early on to pay attention to specific aspects of motion events and experience" (1996: 94). Seen in this way, younger children of both Chinese and English may have the same pre-linguistic potential for conceptualising Path as the most salient and central ingredient for motion. However, as they develop linguistically, they are selectively prompted by the structure of their respective input to view Manner and/or Path as more salient (see also Choi and Bowerman 1991).

Our results are both congruent and discordant with previous investigations of linguistic relativity in the spatial domain (e.g., Gennari et al. 2002; Papafragou et al. 2002; Soroli and Hickmann 2010). In particular, our finding iii., echoes with those obtained by Hohenstein (2005), who revealed that linguistic and non-linguistic performance were associated with each other. English 7-year-olds fixated on videos matching the Manner (versus Path) of a target video more often than did Spanish 7-year-olds and younger children of 3.5 years in both languages, thus reflecting a developmental pattern that has already been attested in language acquisition. In fact, our results reinforce Hohenstein's (2005) findings in two aspects. Firstly, in Hohenstein's (2005) study, no shadowing condition was provided, therefore, it is possible that participants had verbalised motion scenes while they watched the videos, and may have gone on to use such information – particularly the salience of the verb – when choosing the matching alternate videos. However, our findings were obtained in a truly 'non-linguistic' situation. Participants were required to perform an age-appropriate verbal distractor task: counting backwards for adults and counting forwards for the 8-year-olds, which has contributed to eliminating any possible linguistic influence on non-linguistic thought during similarity judgments. Secondly, Hohenstein (2005) examined two languages with contrasting typological properties (English versus Spanish), whereas in the present study the two languages in question are at least partially similar. In some sense, finding iii. is of particularly striking significance in revealing that the effects of language are strong enough to lead to variations in the pattern of non-linguistic thought even with minimal differences between languages.

## 8 Conclusions

On the whole, our results seem to lend further support to the hypothesis of linguistic relativity: before acquiring the linguistic differences in question, the 3-year-olds were not different in their similarity judgments of caused motion events (both groups were Path-oriented), but after children became accustomed to using the spatial language in ways typical to their native language, the pattern of their non-linguistic thought showed language-biased variations: English children shifted to a more Manner-oriented perspective whilst Chinese children developed no particular preference for either Manner or Path dimension. These observations suggest an overarching effect of spatial language use on the maturity of spatial cognition, thus shedding fresh light on the relationship between language and thought in general.

Much ground has yet to be covered in the study of language, cognition and motion events. There are several important lines of research that future studies may follow. First of all, representative verb-framed languages such as Spanish can be included in future investigations of motion event cognition in order to reveal whether speakers of the verb-framed group demonstrate different preferences and/or reaction patterns in categorization tasks such as similarity judgment. Given the current set of research findings, it seems plausible to hypothesize that in terms of overt choices, speakers of verb-framed languages may resemble their counterparts from satellite- and equipollently-framed groups in predominately opting for the Path-match. This predication is based on the observation that there are robust variations in terms of framing properties within a given language and the attested crosslinguistic differences in motion event encoding tend to be probabilistic rather than absolute, thus rendering it unlikely to incur categorical differences in behaviour. In terms of response latency, however, we expect that speakers of verb-framed languages may possibly demonstrate a pattern in which their response time to the Path-match will be significantly shorter than that to the Manner-match. That is, they are potentially more oriented to the Path dimension and may have greater efficiency in the subconscious processing of Path information.

Another venue of future research is to look at motion event cognition amongst second language (L2) learners. According to Slobin (1996), one's native language is not a neutral coding system of an objective reality; Instead, it trains its speakers, from childhood, to pay habitual attention to specific dimensions of experience that are already enshrined in grammatical categories. In this light, if a particular language instantiates a special way of thinking, and differences in linguistic structure foster variations in cognitive patterns, then what would

happen to speakers who have a command of two or more languages? Will they look at the world differently than monolingual speakers? To what extent can they remould their cognitive disposition as a result of additional language learning? What is the nature of the dynamic relationship between progress in L2 acquisition and the shifting cognitive state of an L2 speaker (as discussed in Bylund and Athanasopoulos 2015: 953–954, and Ji 2017: 1–2)?

Further, in order to fully account for the intricacies of crosslinguistic differences in semantic-conceptual structures in motion event cognition, more varied types of data are needed, such as those concerning the type of motion under investigation (spontaneous or voluntary motion vs. caused motion; see Ji and Hohenstein 2017b) and those of specific event properties that may exert an influence on speakers’ behaviour (e.g., more Manner-match inducing vs. more Path-match eliciting).

**Acknowledgements:** We would like to thank the National Social Sciences Foundation of China (15BYY063) and the British Academy Postdoctoral Fellowship (PF100022) for their generous support of the research. We extend our thanks to Dr Xiaobei Zheng for her technical support with data analysis.

Appendix A

Sixteen triads of video stimuli in similarity judgment

Item	Target	Manner-match	Path-match
1	Kick balloon up hill	Kick balloon down hill	Throw balloon up hill
2	Pull bag up pyramid	Pull bag into pyramid	Push bag up pyramid
3	Roll ring down dune	Roll ring away from dune	Push ring down dune
4	Pull hay up ladder	Pull hay around ladder	Roll hay up ladder
5	Kick ball into puddle	Kick ball out of puddle	Roll ball into puddle
6	Kick ball across slide	Kick ball down slide	Push ball across slide
7	Push boat out of lake	Push boat towards lake	Pull boat out of lake
8	Drag toy car across ice	Drag toy car around ice	Slide toy car across ice
9	Push log towards campfire	Push log away from campfire	Roll log towards campfire
10	Roll sack towards escalator	Roll sack up escalator	Slide sack towards escalator
11	Pull suitcase away from tent	Pull suitcase into tent	Slide suitcase away from tent
12	Pull tyre towards rails	Pull tyre along rails	Throw tyre towards rails

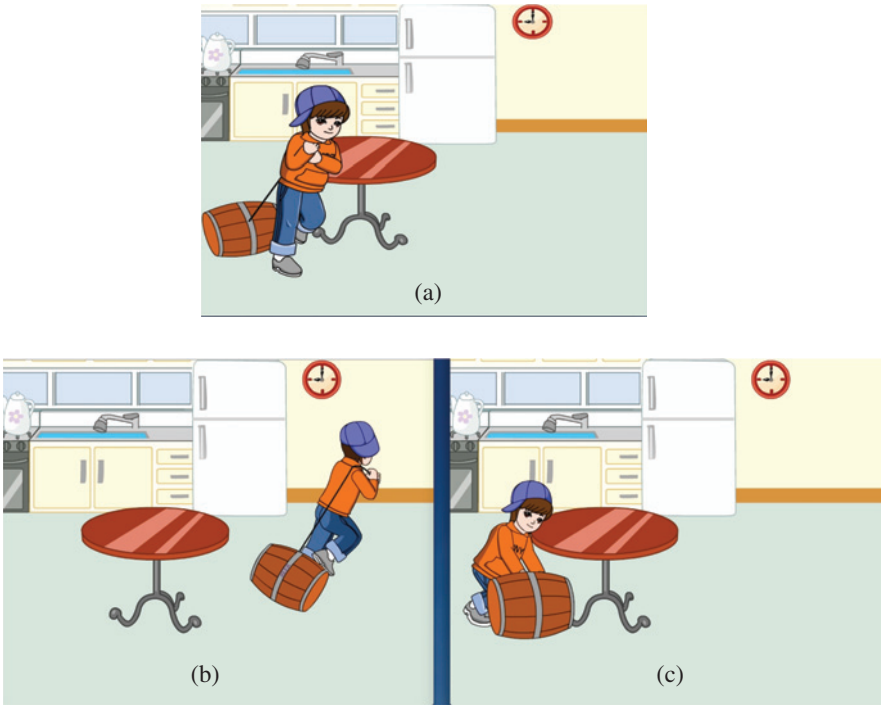
(continued)

(continued)

Item	Target	Manner-match	Path-match
13	Roll basketball along chairs	Roll basketball around chairs	Throw basketball along chairs
14	Push balloon around slide	Push balloon up slide	Kick balloon around slide
15	Pull box along tunnel	Pull box out of tunnel	Kick box along tunnel
16	Drag barrel around table	Drag barrel away from table	Roll barrel around table

Appendix B

An illustration of the video stimuli used in the experiment: triad 16



16: (a) Dragging a barrel around table. (b) Dragging a barrel *away from* table. (c) *Rolling* a barrel around table.

Audio stimuli accompanying the video clips were:

Target: This is sixteen.

Alternates: Which one is most like sixteen?

## References

- Allen, S., A. Özyürek, S. Kita, A. Brown, R. Furman, T. Ishizuka & M. Fujii. 2007. Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition* 102(1). 16–48.
- Ameka, F. K. & J. Essegbey. 2013. Serialising languages: Satellite-framed, verb-framed or neither. *Ghana Journal of Linguistics* 2(1). 19–38.
- Athanasopoulos, P. & E. Bylund. 2013. Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science* 37. 286–309.
- Beavers, J., B. Levin & S. Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics* 46. 331–377.
- Berman, R. & D. I. Slobin. 1994. *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum.
- Berthele, R. 2004. The typology of motion and posture verbs: A variationist account. In B. Kortmann (ed.), *Dialectology meets typology*, 93–126. Berlin: Mouton de Gruyter.
- Berthele, R. 2013. Disentangling manner from path: Evidence from varieties of German and Romance. In J. Goschler & A. Stefanowitsch (eds.), *Variation and change in the encoding of motion events*, 55–76. Amsterdam: John Benjamins.
- Berthele, R. 2015. Convergence in the domains of static spatial relations and events of putting and taking. Evidence from bilingual speakers of Romansh and German. *International Journal of Bilingual Education and Bilingualism* 18(5). 624–642.
- Blomberg, J. & J. Zlatev. 2009. Linguistic relativity, mediation and the categorization of motion. In J. Zlatev, M. Andrén, M. Johansson Falck & C. Lundmark (eds.), *Studies in language and cognition*, 46–61. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Bohnemeyer, J., S. Eisenbeiss & B. Narasimhan. 2004. Manner and path in non-linguistic cognition. Paper presented at the International Conference on Language, Culture & Mind, University of Portsmouth, 18–20 July.
- Boroditsky, L. 2000. Metaphoric structuring: Understanding time through spatial metaphors. *Cognition* 75(1). 1–28.
- Boroditsky, L. 2001. Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology* 43. 1–22.
- Boroditsky, L. & A. Gaby. 2010. Remembrances of times east: Absolute spatial representations of time in an Australian aboriginal community. *Psychological Science* 21. 1635–1639.
- Bowerman, M. 1982. Starting to talk worse: Clues to language acquisition from children's late speech errors. In S. Strauss (ed.), *U-shaped behavioral growth*, 101–146. New York: Academic Press.
- Bowerman, M. 1988. The child's expression of meaning: Expanding relationships among lexicon, syntax, and morphology. In M. B. Franklin & S. S. Barten (eds.), *Child language: A reader*, 106–117. New York & Oxford: Oxford University Press.
- Bowerman, M. 1996. The origins of children's spatial semantic categories: Cognitive versus linguistic determinants. In J. Gumperz & S. Levinson (eds.), *Rethinking linguistic relativity*, 145–176. Cambridge: Cambridge University Press.
- Bowerman, M. 1999. Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, M. Peterson, L. Nadel & M. Garrett (eds.), *Language and space*, 385–436. Cambridge, MA: MIT Press.

- Bowerman, M. & S. Choi. 2001. Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (eds.), *Language acquisition and conceptual development*, 475–511. Cambridge: Cambridge University Press.
- Bylund, E. & P. Athanasopoulos. 2015. Introduction: Cognition, motion events, and SLA. *The Modern Language Journal* 99(S1). 1–13.
- Cadierno, T., I. Ibarretxe-Antuñano & A. Hijazo-Gascón. 2016. Semantic categorization of placement verbs in L1 and L2 Danish and Spanish. *Language Learning* 66(1). 191–223.
- Chen, J. 2008. *The acquisition of verb compounding in Mandarin Chinese*. Amsterdam: Vrije Universiteit Amsterdam unpublished dissertation.
- Chen, L. 2005. *The acquisition and use of motion event expressions in Chinese*. Lafayette, LA: University of Louisiana unpublished dissertation.
- Chen, L. & J. Guo. 2009. Motion events in Chinese novels: Evidence for an equivalently-framed language. *Journal of Pragmatics* 41. 1749–1766.
- Choi, S. & M. Bowerman. 1991. Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition* 41. 83–121.
- Chu, C. 2009. Path of motion: Conceptual structure and representation in Chinese. In J. Xing (ed.), *Studies of Chinese linguistics: Functional approaches*, 65–73. Hong Kong: Hong Kong University Press.
- Dittmar, M., K. Abbot-Smith, E. Lieven & M. Tomasello. 2008. Young German children's early syntactic competence: A preferential looking study. *Developmental Science* 11(4). 575–582.
- Durst-Andersen, P. 2011. *Linguistic supertypes: A cognitive-semiotic theory of human communication*. Berlin: Mouton de Gruyter.
- Fernald, A., R. Zangl, A. L. Portillo & V. A. Marchman. 2008. Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernandez & H. Clahsen (eds.), *Developmental psycholinguistics: On-line methods in children's language processing*, 97–135. Amsterdam: John Benjamins.
- Filipović, L. 2007. On the nature of lexicalization patterns: A cross-linguistic enquiry. In N. Delbecque & B. Cornillie (eds.), *On interpreting construction schemas*, 307–330. Berlin & New York: Mouton de Gruyter.
- Filipović, L. 2013. Constructing causation in language and memory: Implications for access to justice in multilingual interactions. *International Journal of Speech, Language and the Law* 20. 1–19.
- Filipović, L. & K. Jaszczołt (eds.). 2012. *Space and time in languages and cultures: Linguistic diversity*. Amsterdam: John Benjamins.
- Finkbeiner, M., J. Nicol, D. Greth & K. Nakamura. 2002. The role of language in memory for actions. *Journal of Psycholinguistics Research* 31(5). 447–457.
- Flecken, M. 2011. Event conceptualization by early Dutch-German bilinguals: Insights from linguistic and eye-tracking data. *Bilingualism: Language and Cognition* 14(1). 61–77.
- Flecken, M., C. Von Stutterheim & M. Carroll. 2014. Grammatical aspect influences motion event perception: Findings from a cross-linguistic nonverbal recognition task. *Language and Cognition* 6. 45–78.
- Gao, H. 2001. Notions of motion and contact for physical contact verbs. In A. Holmer, J. Svantesson & A. Viberg (eds.), *Proceedings of the 18th Scandinavian conference of linguistics*, Vol. 2, 193–209. Lund: University of Lund Department of Linguistics.
- Gennari, S. P., S. Sloman, B. Malt & T. Fitch. 2002. Motion events in language and cognition. *Cognition* 83. 49–79.

- Goschler, J. & A. Stefanowitsch (eds.). 2013. *Variation and change in the encoding of motion events*. Amsterdam: John Benjamins.
- Gullberg, M. 2011. Language-specific encoding of placement events in gestures. In J. Bohnemeyer & E. Pederson (eds.), *Event representation in language and cognition*, 166–188. Cambridge, MA: Cambridge University Publishing.
- Gullberg, M. & B. Narasimham. 2010. What gestures reveal about how semantic distinctions develop in Dutch children's placement verbs. *Cognitive Linguistics* 21. 239–262.
- Hickmann, M. 2003. *Children's discourse: Person, space and time across languages*. Cambridge: Cambridge University Press.
- Hickmann, M. 2006. The relativity of motion in first language acquisition. In M. Hickmann & S. Robert (eds.), *Space in languages: Linguistic systems and cognitive categories*, 281–308. Amsterdam: John Benjamins.
- Hickmann, M., H. Hendriks & C. Champaud. 2009. Typological constraints on motion in French and English child language. In J. Guo, E. Lieven, S. Ervin-Tripp, N. Budwig, Ş. Özçalışkan & K. Nakamura (eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*, 209–224. New York: Psychology Press.
- Hohenstein, J. 2005. Language-related motion event similarities in English- and Spanish-speaking children. *Journal of Cognition and Development* 6. 403–425.
- Hohenstein, J., L. Naigles & A. Eisenberg. 2004. Keeping verb acquisition in motion: A comparison of English and Spanish. In G. Hall & S. Waxman (eds.), *Weaving a lexicon*, 569–602. Cambridge, MA: MIT Press.
- Iacobini, C. & F. Masini. 2006. The emergence of verb-particle constructions in Italian locative and actional meanings. *Morphology* 16. 155–188.
- Ibarretxe-Antuñano, I. 2009. Path salience in motion events. In J. Guo, E. Lieven, S. Ervin-Tripp, N. Budwig, Ş. Özçalışkan & K. Nakamura (eds.), *Crosslinguistic approach to the psychology of languages: Research in the tradition of Dan Isaac Slobin*, 403–414. New York: Psychology Press.
- Ibarretxe-Antuñano, I. 2012. Placement and removal events in Basque and Spanish. In A. Kopecka & B. Narasimhan (eds.), *Events of “putting” and “taking”: A crosslinguistic perspective*, 123–144. Amsterdam & Philadelphia: John Benjamins.
- Jackendoff, R. 1996. The architecture of the linguistic-spatial interface. In P. Bloom, M. Peterson, L. Nadel & M. Garrett (eds.), *Language and space*, 1–30. Cambridge, MA: MIT Press.
- Ji, Y. 2017. Motion event similarity judgments in one or two languages: An exploration of monolingual speakers of English and Chinese vs. L2 learners of English. *Frontiers in Psychology – Language Sciences* 8. 909. doi: 10.3389/fpsyg.2017.00909 (accessed 8 June 2017).
- Ji, Y., H. Hendriks & M. Hickmann 2011a. Children's expression of voluntary motion events in English and Chinese. *Journal of Foreign Languages* 34(4). 2–20.
- Ji, Y., H. Hendriks & M. Hickmann. 2011b. How children express caused motion events in Chinese and English: Universal and language-specific influences. *Lingua* 121. 1796–1819.
- Ji, Y., H. Hendriks & M. Hickmann. 2011c. The expression of caused motion events in Chinese and in English: Some typological issues. *Linguistics* 49(5). 1041–1076.
- Ji, Y. & J. Hohenstein. 2017a. Replication Data for: English and Chinese children's motion event similarity judgments, doi:10.18710/AAZVJH, DataverseNO, V1 (accessed 9 August 2017).
- Ji, Y. & J. Hohenstein. 2017b. Stimulus Triads for: English and Chinese children's motion event similarity judgments, doi:10.18710/V8CAOU, DataverseNO, V1 (accessed 9 August 2017).



- Ji, Y. & J. Hohenstein. 2017c. Conceptualising voluntary motion events beyond language use: A comparison of English and Chinese speakers' similarity judgments. *Lingua* 195. 57–71.
- Kang, J. 1999. The composition of the perfective aspect in Mandarin Chinese. Boston, MA: Boston University unpublished dissertation.
- Kopecka, A. 2006. The semantic structure of motion verbs in French: Typological perspectives. In M. Hickmann & S. Robert (eds.), *Space in languages: Linguistic systems and cognitive categories*, 83–101. Amsterdam: John Benjamins.
- Kopecka, A. & B. Narasimhan (eds.). 2012. *Events of putting and taking: A crosslinguistic perspective*. Amsterdam: John Benjamins.
- Landau, B. & L. Lakusta. 2006. Spatial language and spatial representation: Autonomy and interaction. In M. Hickmann & S. Roberts (eds.), *Space in languages: Linguistic systems and cognitive categories* (Typological Studies in Language 66), 309–333. Amsterdam: John Benjamin.
- Levinson, S. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.
- Li, C. N. & S. A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
- Li, Y. F. 1990. On V–V compounds in Mandarin. *Natural Language and Linguistic Theory* 8. 177–207.
- Lucy, J. 1992. *Grammatical categories and cognition*. Cambridge: Cambridge University Press.
- Lupyan, G. 2012. Linguistically modulated perception and cognition: The label–feedback hypothesis. *Frontiers in Psychology* 3. 54. doi:10.3389/fpsyg.2012.00054 (accessed 12 October 2016).
- Mandler, J. M. 1992. How to build a baby II: Conceptual primitives. *Psychological Review* 99. 587–604.
- McDonald, E. 1995. Completive verb compounds in modern Chinese: A new look at an old problem. *Journal of Chinese Linguistics* 22. 317–362.
- McWhorter, J. H. 2014. *The language hoax. Why the world looks the same in any language*. Oxford: Oxford University Press.
- Papafragou, A., J. Hulbert & J. Trueswell. 2008. Does language guide event perception? Evidence from eye movements. *Cognition* 108(1). 155–184.
- Papafragou, A., C. Massey & L. Gleitman. 2002. Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition* 84. 189–219.
- Pavlenko, A. & M. Volynsky. 2015. Motion encoding in Russian and English: Moving beyond Talmy's typology. *The Modern Language Journal* 99(S1). 32–48.
- Pederson, E., E. Danziger, D. Wilkins, S. C. Levinson, S. Kita & G. Senft. 1998. Semantic typology and spatial conceptualization. *Language* 74(3). 557–589.
- Pulverman, R., L. Song, K. Hirsh-Pasek & S. M. Pruden. 2013. Preverbal infants' attention to manner and path: Foundations for learning relational terms. *Child Development* 84(1). 241–252.
- Skopeteas, S. 2008. Encoding spatial relations: Language typology and diachronic change in Greek. *STUF – Language Typology and Universals* 61(1). 54–66.
- Slobin, D. I. 1996. From “thought and language” to “thinking to speaking”. In J. J. Gumperz & S. C. Levinson (eds.), *Rethinking linguistic relativity*, 70–96. Cambridge: Cambridge University Press.
- Slobin, D. I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (eds.), *Relating events in narrative, Vol. 2: Typological and contextual perspectives*, 219–257. Mahwah, NJ: Lawrence Erlbaum.



- Soroli, E. & M. Hickmann. 2010. Language and spatial representations in French and in English: Evidence from eye-movements. In G. Marotta, A. Lenci, L. Meini & F. Rovai (eds.), *Space in language*, 581–597. Pisa : Editrice Testi Scientifici.
- Spelke, E. S. 2003. What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (eds.), *Language in mind: Advances in the investigation of language and thought*, 277–311. Cambridge, MA: MIT Press.
- Tai, J. 2003. Cognitive relativism: The resultative construction in Chinese. *Language and Linguistics* 4(2). 302–316.
- Talmy, L. 1985. Lexicalization patterns: Semantic structure in lexical form. In T. Shopen (ed.), *Language typology and syntactic description* Vol. 3, 36–149. Cambridge: Cambridge University Press.
- Talmy, L. 2000. *Toward a cognitive semantics, Vol. 2: Typology and process in concept structuring*. Cambridge, MA: MIT Press.
- Talmy, L. 2009. Main verb properties and equipollent framing. In J. Guo, E. Lieven, S. Ervin-Tripp, N. Budwig, Ş. Özçalışkan & K. Nakamura (eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*, 389–402. New York: Psychology Press.
- Von Stutterheim, C., M. Andermann, M. Carroll, M. Flecken & B. Schmiedtová. 2012. How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics* 50(4). 833–867.
- Wolff, P. 2003. Direct causation in the linguistic coding and individuation of causal events. *Cognition* 88(1). 1–48.
- Wolff, P. & K. Holmes. 2011. Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science* 2. 253–265.
- Zlatev, J. 2011. From cognitive to integral linguistics and back again. *Intellectica* 56. 125–147.
- Zlatev, J. & J. Blomberg. 2015. Language may indeed influence thought. *Frontiers in Psychology* 6. 1631. doi: 10.3389/fpsyg.2015.01631 (accessed 12 October 2016).
- Zlatev, J. & P. Yangklang. 2004. A third way to travel: The place of Thai in motion event typology. In S. Strömquist & L. Verhoeven (eds.), *Relating events in narrative: Cross-linguistic and cross-contextual perspective*, 159–190. Mahwah, NJ: Lawrence Erlbaum.
- Zou, K. 1994. Directional verb compounds in Chinese. In K. Beals, J. Denton, B. Knippen, L. Melnar, H. Suzuki & E. Zeinfeld (eds.), *CLS 30-I: Papers from the thirtieth regional meeting of the Chicago Linguistic Society*, 443–457. Chicago, Illinois: The Chicago Linguistic Society.